The Bayesian approach in the valuation – a strategy to handle markets with low purchasing prices?

Alexandra WEITKAMP and Hamza ALKHATIB, Germany

Key words: valuation, Bayes analysis; regression analysis; low purchasing prices

SUMMARY

The real estate and finance crisis has shown the importance of real estate valuation. The real estate value has to satisfy high objective requirements. Besides, the German jurisdiction demands a maximum dispersion of ± 20 % of the market value.

The expertise of real estate valuers can be regarded as very accurate. However, in the technical part of the valuation, especially in the regression analysis – the expertise remains unused. The expertise influences the results significantly by establishing the model. The following “pure” computation is normally done by software routines, which do not need an intervention by experts. Here, a Bayesian approach is introduced to integrate experts’ knowledge in data-driven models, such as the multiple linear regression analysis.

In this Bayesian analysis, we will focus only on one and two family houses. The dependent variable is the price per living space. The independent variables are the area of lot, the standard land value, the age, the living space and the equipping standard. First, the expertise is obtained from interviews with valuation experts. They estimate the changes of the market value depending on variations in the parameters of the valuation object. This prior information is summarized in a sample and treated within a multiple linear regression. The resulting regression function is expressed as a continuous normal density function, which would be used as prior density in the Bayesian approach. The dependent variables (purchasing prices) are taken to derive the likelihood function, which is also a normal density function. By means of Bayes's theorem, we obtain the posterior density function which is proportional to the prior density multiplied by the likelihood function. Based on this posterior density, improved mean values of the regression coefficients and their more certain accuracy can be derived.

In order to simulate the market with low purchasing prices, the data is reduced by particular cases in the next steps. The selection is done by special criteria: cases in the edges of the sample, in the center of the data, spatial selection (e. g. special municipalities) etc. In these locations, the experts normally determine the market value from their expertise with a look on the few purchasing prices. Based on the Bayesian approach, they can use both: data and knowledge in a comprehensive model. The aim hereby is to establish a procedure, which allows using small sample data in combination with the experts’ knowledge, and provides statistically certain results.

The Bayesian approach in the valuation – a strategy to handle markets with low purchasing prices?

Alexandra WEITKAMP and Hamza ALKHATIB, Germany

1. MOTIVATION

It is necessary that the real estate value has to satisfy high objective requirements. The German jurisdiction demands a maximum dispersion of ± 20 % of the market value (BVerfG, Urt. v. 07.11.2006 1 BvL 10/02). The conformance to this requirement is easy, if there is enough useful data. But the expert committees have to derive valuation data like standard land values e. g. for building land in all locations in Germany even if there are only a few purchases. In these cases, data-driven models as the classical regression analysis normally fail. The experts committees have to rely on their experiences and estimate the values by their knowledge.

Anyhow, the good expertise knowledge is not integrated in the data-driven models in the past. It is only needed to establish the model. The expertise knowledge is not used to increase the model and the uncertainty. In Alkhatib and Weitkamp (2012) and Weitkamp and Alkhatib (2012) a new approach using Bayes interference is presented, which combines the expertise knowledge with data-driven models on basis of the Bayes theorem. Now, it will be shown, that the approach could be used in markets with low purchasing prices and which improvements the approach will entail.

2. BAYESIAN APPROACH IN VALUATION

Due to the lack of space, only a small insight into theoretical background will be given. The reader refered to Alkhatib and Weitkamp (2012) for more details on the Bayesian approach. A closer view on expert knowledge and a first application is given in Weitkamp and Alkhatib (2012).

In the valuation practice, the regression analysis is established for analyzing purchase data. Especially for the valuation of standard land values, it is the preferred method. A detailed motivation and discussion of the regression model can be found in any standard statistics textbook (e. g. Ziegenbein 1977, Brückner 1976, Fahrmeir et al. 2009, Urban and Mayerl 2006). Briefly, the linear regression model studies the relationship between a dependent variable **y** and *k* independent variables $\mathbf{X}_1 \ldots \mathbf{X}_k$, summarized in the matrix **X**, according to:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \text{Eq. 1}$$

The regression coefficients $\boldsymbol{\beta}$ are estimated according to the method of least squares with $\boldsymbol{\varepsilon}$ as residuals (e. g. Brückner 1976 or Fahrmeir et al. 2009).

2.1 Modeling the experts' knowledge for a Bayesian approach

The expert decides, which independent variables $\mathbf{X}_i$ should be analyzed regarding to an influence on the dependent variable **y**. Now, the expertise should be used within the model. But what does a good expert or an appropriate expertise makes up? The valuation experts are members of the professions of surveyors, economists, civil engineers / architects, lawyers or estate agents. Therefore, they constitute a different technical expertise. In Germany, a statutory license is not necessary; however, the expert has to demonstrate competence and sufficient experience (Gondring 2004: 943 f.). Ziegenbein (2010) postulates that valuers should have special knowledge about real estate values by education and experience. They should act professionally, fact-based and objective. Experts should work independently and neutrally. They should be not financially or personally dependent on the principal (Ziegenbein 2009). Experts in valuation usually have a great experience. In Germany, valuers often prove it by certification, e. g. according to ISO/IEC 17024 (Deutsches Institut für Normung 2009). Many valuation methods adopt this knowledge. The less data (purchasing prices) there is, the more challenging an appropriate valuation exists. Often an assessment can only be achieved by the fact, that the knowledge is transferred from other submarkets to the present evaluation object (Ziegenbein 2009). Reuter (2006) uses experts' expertise for the method of intersubjective price comparison in markets with low purchasing prices.

But the expertise remains in the purely technical part of the regression analysis unused. It is only used to prepare the model. In the Bayesian theory, it is possible to use previous knowledge about the independent variables in data-driven models. In this approach, the prior knowledge is integrated by using probability density functions (PDF) – so called prior densities (Alkhatib and Weitkamp 2012).

Probability density functions are related with the so called probability distributions. Probability distributions of a random variable indicate the probability of the values of this random variable. The probability distribution could be presented as:

$$F(x) = P(X \leq x). \qquad \text{Eq. 2}$$

X is the random variable, x is the realisation of X and P is the probability (Koch 2007). Probability distributions give probabilities for realizations of a random variable. They show how the distribution of probabilities to the possible random results (realizations), especially they show the possible values of the random variable.

Otherwise, probability density functions describe these probability distributions. The integration of probability density functions over an interval [a,b] gives the probability, that a random variable with this density takes a value between a and b:

$$P(a < X < b) = F(b) - F(a) = \int_a^b f(t)dt. \qquad \text{Eq. 3}$$

The probability density function is the derivative of the distribution function by the random variables (Sachs and Hedderich 2006: 145 ff.). The prior knowledge is modeled in a probability density function and processed in the Bayesisan approach. If the functional relationship of the factors is linear and no prior knowledge exists, the results of the Bayesian approach and the multiple linear regression analysis are identical (Alkhatib and Weitkamp 2012).

It should be noted that the Bayesian approach should be applied only when the "quality" of the valuers is ensured. The experts' knowledge affects the results; accurate experts' knowledge will lead to more certain outcomes. A wrong judgement will cause a negative effect on the resulting output (regression function). Maybe the quality could be assured by e. g. certification, but can also be provided by other evidence. Specifically, care should be taken to ensure that the expertise is based on several experts. They should assess the same facts redundantly to detect possible outliers in the estimation. Also, it must be guaranteed, that the experts act independently to ensure their autonomous judgment (Weitkamp and Alkhatib 2012).

2.2 Bayesian Enhancements

Further on, the Bayesian approach is shortly introduced. In the Bayesian interference, we are interested in learning about the regression coefficients $\boldsymbol{\beta}$ based on the data $\mathbf{y}$. To achieve the above mentioned goal, we use the Bayes' rule:

$$P(\boldsymbol{\beta} \mid \mathbf{y}) \propto P(\boldsymbol{\beta}) P(\mathbf{y} \mid \boldsymbol{\beta}) \qquad \text{Eq. 4}$$

The term $P(\boldsymbol{\beta}|\mathbf{y})$ is the posterior density. The posterior density contains the knowledge on the parameters $\boldsymbol{\beta}$ by given data $\mathbf{y}$. $P(\mathbf{y}|\boldsymbol{\beta})$ is called Likelihood function and summarizes the given data (purchases). The term $P(\boldsymbol{\beta})$ contains the prior information. A further motivation and discussion of the Bayesian interference can be found in a variety of standard works (e. g. Koch 2007, Kacker and Jones 2003).

The prior information is generated by interviews with valuation experts. A description of the experiment can be found in section 3. Within the interview, the experts have to estimate the impact of variations in the dependent variables on the market value. All of these estimations were used for generating a "new sample" of (experts') data. With this data, a multiple linear regression analysis is done. The result is a regression function with an uncertainty budget for the regression coefficients. This regression function is used to derive normally distributed prior densities $P(\boldsymbol{\beta})$. These prior densities are used in the following Bayesian approach as informative prior densities. If there is no previous knowledge about the coefficients, the prior density is assumed to be constant and used as so called non-informative prior densities (Weitkamp and Alkhatib 2012).

The data on purchasing prices is used for the Likelihood function. The data is analyzed by multiple linear regression analysis and the regression coefficients $\boldsymbol{\beta}$ are estimated. Dependent variable $\mathbf{y}$ is the variable "price per living space" in €/m². The independent variables $\mathbf{X}$ are "area of lot" [m²], "standard land value" [€/m²], building year [a], "living space" [m²] and "equipping standard" [without unit] (Weitkamp and Alkhatib 2012).

The result of the Bayesian approach is the posterior density function $P(\boldsymbol{\beta}|\mathbf{y})$. In the case of multiple linear regression analysis, this function is analytically solvable. This is due to the assumption of normal distribution and linearity. For all solutions, the posteriori regression coefficients $\bar{\boldsymbol{\beta}}$, the uncertainty (cofactor matrix $\bar{\mathbf{V}}$) and especially the confidence intervals (Highest Posterior Density Intervals / HPDI) can be derived. In the case of informative prior information, they can be presented (see Tab. 1) with $\underline{\boldsymbol{\beta}}$, $\underline{\mathbf{V}}$ as prior regression coefficient resp. cofactor matrix.

Tab. 1: Comparison between classical and Bayesian approach concerning regression coefficients and cofactor matrix.

Classical Regression Analysis		*Bayesian Parameter Estimation*	
$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$	Eq. 5	$\bar{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \underline{\mathbf{V}}^{-1})^{-1}(\mathbf{X}'\mathbf{y} + \underline{\mathbf{V}}^{-1}\underline{\boldsymbol{\beta}})$	Eq. 6
$\mathbf{V} = \mathbf{Q}_{\boldsymbol{\beta\beta}} = (\mathbf{X}'\mathbf{X})^{-1}$	Eq. 7	$\bar{\mathbf{V}} = (\mathbf{X}'\mathbf{X} + \underline{\mathbf{V}}^{-1})^{-1}$	Eq. 8

Having a closer look at these equation, it must be recognized, that the posterior parameters (overlined) are estimated on prior parts (red and underlined) and data parts (blue). In a classic regression analysis (Eq. 5 and Eq. 7), the prior parts (red) do not exist like in the Bayesian approach (Eq. 6 and Eq. 8). The outstanding result of the posterior density is a more precise estimate of the regression coefficients. In addition, the uncertainty and confidence intervals, respectively, are determined. The corresponding mathematical procedure for deriving the priority densities, the likelihood function and posterior densities has been developed and described in detail in Alkhatib and Weitkamp (2012).

3. APPLICATION OF THE BAYESIAN APPROACH IN A CHOSEN SUBMARKET

The Bayesian approach is useful to integrate prior knowledge. This knowledge can be derived from former data analyses or from experts by interview. Here, the knowledge was recovered by interviewing valuers of the expert committee.

3.1 Submarket und Interview

As study area the region of Osnabrueck was selected as spatial submarket. Osnabrueck is located in the west of Lower Saxony. The area was chosen for its good data set. As statutory task, the committee of valuation experts records purchases in the data collection on purchasing prices. In addition to the statutory task, they are labeling purchases, where they do an appraisal. For this reason, purchases with a former appraisal can be selected like it is needed in the Bayesian approach.

Free-standing one and two family houses as well as semi-detached and row houses are chosen as functional submarket. All purchase cases originate from ordinary course of business without personal relationships.

For modeling the prior information, 15 appraisals are used: 10 appraisals are assessing free-standing one family houses, five are assessing row-houses. Six appraisals are situated in the city of Osnabrueck. The experts were chosen from designated estate agents, bankers and official appraisers. Each expert has to appraise five objects. The expert should establish the market value for each object. He has to appraise only objects which are situated in their common working environment. The expert obtains all necessary information to appraise the market value. He has to estimate the change of the market value for three variations in the value-relevant factors. These are the independent variables “area of lot”, “standard land value”, “building year”, “living space” and “equipping standard”. They know the market of the objects and could estimate the effect of the variations on the market value of the object. The variations are taken within realistic borders for the special valuation object. For each object, three variations are taken. It was ensured that the sign of the variations differs within one object. Also the variation of each independent variable differs in the sign. For each object, appraisals of 3 – 5 experts are available as result. Each expert appraises the market value and the effect of three variations of each object. Also, the original appraisal is used further on. The experts should appraise as usual: If they usually calculate the market value, they should calculate it in the experiment, as well. If they usually estimate form their expert knowledge, they should do this also. As a result, 270 new non-real “purchases” are generated.

These new data set is used to obtain the prior information and based on that, we are able to build the prior density (Weitkamp and Alkhatib 2012).

3.2 Overview on the results for the whole sample

In this approach, the sample of purchases is analyzed as usually in a regression analysis. The following regression model is established:

$$\mathbf{y} = \mathbf{X\beta} = [\mathbf{1} \quad \mathbf{X}_2 \quad \mathbf{X}_3 \quad \dots \quad \mathbf{X}_6] \cdot [\beta_1 \quad \beta_2 \quad \beta_3 \quad \dots \quad \beta_6]'. \qquad \text{Eq. 9}$$

The dependent variable **y** contains the prices per living space. The independent variables $\mathbf{X}_2$ … $\mathbf{X}_6$ as vector includes the *n* data of the area of lot ($\mathbf{X}_2$), the area of living space ($\mathbf{X}_3$), the standard land value ($\mathbf{X}_4$), the age ($\mathbf{X}_5$), and the equipping standard (X_6). Based on $n = 1000$ purchases, the regression function could be estimated as:

Price per living space = 1078,37 + 0,39 Area of lot – 3,56 Area of living space + 3,95 Standard land value – 10,05 Age + 66,96 Equipping standard. Eq. 10

The likelihood function can be derived from Eq. 10, which summarizes the information of the real data (purchases). The informative prior density is derived from the interviews made by experts. Besides the regression analysis of the data set, a regression analysis is done to obtain the prior data. Therefore, the data set of the experts' appraisals is used. The data is handled like a common data set and with the same approach like the data of purchases.

Tab. 2 shows that values for the mean of the regression coefficients in both cases, the informative and non-informative prior which are quite similar. In addition, we see that the most mean values of the regression coefficients for informative prior lie between the prior mean and the posterior mean of non-informative prior. Please note that the LS-estimate of the regression coefficients and the posterior derived from the non-informative prior are identical in this study. Only the posterior (informative) mean of β_3 does not lie between prior mean and posterior mean using non-informative prior (Weitkamp and Alkhatib 2012).

Tab. 2: Prior and posterior estimate for the regression coefficients (Weitkamp and Alkhatib 2012).

		Prior	Posterior using	
		Informative	non-informative prior	informative prior
(Intercept)	β_1	668,45	1078,37	965,77
Area of lot	β_2	0,42	0,39	0,40
Area of living space	β_3	-2,17	-3,56	-3,20
Standard land value	β_4	3,60	3,95	3,96
Age	β_5	-7,23	-10,05	-9,81
Equipping standard	β_6	129,04	66,96	82,04

Besides, the confidence intervals are computed for the posterior regression coefficients using non-informative and informative prior. It is obvious, that the confidence intervals of the non-informative prior are more expanded than those of the informative prior. It could be generally noticed, that the confidence regions in case of informative prior are narrower than those in case of non- informative prior. As interpretation, there is clear evidence, that the experts' knowledge improve the results (Weitkamp and Alkhatib 2012).

If the Bayesian approach is used in cases with sufficient data, the main benefit of the informative Bayesian approach is more the improvement of the certainty domain of the estimated regression coefficients than the enhancement of the regression function itself (Weitkamp and Alkhatib 2012).

4. SIMULATION OF MARKETS WITH LOW PURCHASING PRICES

As we see before, in cases of an appropriate number of purchases the Bayesian approach leads to an improvement of the uncertainty rather than the coefficients itselves. This section deals with the question, in which cases an improvement of the estimated regression coefficients based on the Bayesian approach can be reached.

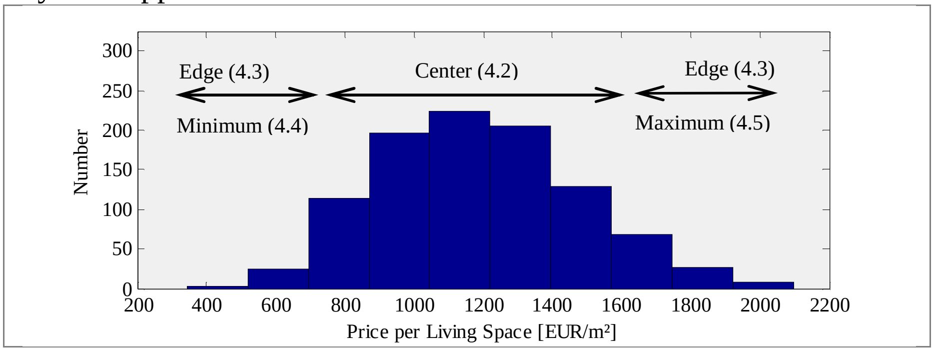

Fig. 1: Methodical reductions of the original sample.

In Weitkamp and Alkhatib (2012), the size of the data-set was sufficiently large. In many valuation cases, however, the sample size is smaller. In so called markets with low purchasing prices, the numbers of cases are too low for an adequate regression analysis or any other statistical approach – Ziegenbein (2009) recommends the use of more than 15 comparable purchases per independent variable. If there are not enough sufficient purchases, the results get more unreliable and skewed. Because of the reliable prior knowledge, the weak database could be equalized and the results stabilized.

Therefore, in the next section, the data sample will be reduced in order to simulate purchases low sample size. The different selections are illustrated in Fig. 1 using the histogram of the dependent variable (prices per living space). Starting with a randomized reduction of the (whole) sample to various minor numbers of purchases (sec. 4.1), the sample will be reduced methodically. Then, a reduction is performed to purchases, which have prices per living space near the center of the whole sample (sec. 4.2). In sec. 4.3, a view on the edges will be done which means that low and high prices per living space will be analyzed. Afterwards, low prices per living space (sec. 4.4) as well as high prices per living space (sec. 4.5) will be investigated separately.

In particular, it is interesting to have a look on the reduction to the maxima (high prices per living space): In markets with low purchasing price, the market has collapsed often. Less good objects will not be sellable – these are the objects which would be found in the minima of a sample of purchases. If a transaction of real estates is possible in these areas that would be probably the better ones (with higher prices per living space).

For all studies, the original sample is randomly reduced based on the above mentioned criteria. First, 150 purchases are randomly leaved out for a cross-validation. Afterwards, *n* purchases were taken without replacement in random process. This experiment is repeated 500 times (we denote this as a Monte Carlo run) and the mean of all parameters is computed. Thereby, a general statement can be made for the results. For the reduced sample, the regression function and the confidence intervals of the original sample is compared with the solution of informative prior information and with the solution of non-informative prior information (equal to the classic multiple linear regression analysis). The results are shown in the following figures.

	Original Sample Informative Solution (black)	Reduced Sample Non-Informative Solution (red)	Reduced Sample Informative Solution (blue)
Regression function Confidence Intervals			

Fig. 2: Legend of Fig. 3 to Fig. 8.

Fig. 2 shows the legend of Fig. 3 – Fig. 8. The original sample is plotted in black. Here, the comparison is always done to the informative solution. The reduced sample based on non-informative prior is plotted in red and the Bayesian approach with informative prior is plotted in blue color.
In addition to the change of the regression coefficients, confidence intervals as measure for uncertainty are depicted. In cases of the certain survey of special areas of the sample, a cross-validation is done. Here, the estimate price of living space using non-informative prior and using informative prior, respectively, are compared with real price of living space. Indicator of quality of the estimated dependent variable $\hat{\mathbf{y}}$ related to the true values (represented by in a cross-validation separated *i* purchases y_i) is the root mean squared error (RMSE) (Willmott and Matsuura 2005, Hartung et al. 2009: 125 f.):

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(\hat{y}_i - y_i \right)^2} . \qquad \text{Eq. 11}$$

4.1 Randomized reduction of the sample

In a first study, there are chosen randomly purchases. It was analyzed a choice between 100 purchases to 700 purchases in steps of 100 purchases. Fig. 3 and Fig. 4 present the regression function of a sample of 100 and 400 purchases, respectively. The regression function is illustrated by varying the standard land value in the borders of its occurrence in the sample. The other independent variables are hold in their mean. The results of the other parameters would be alike.

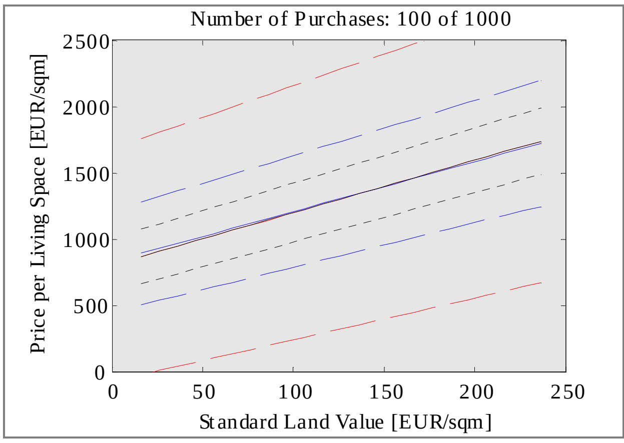

Fig. 3: Regression function for 100 purchases.

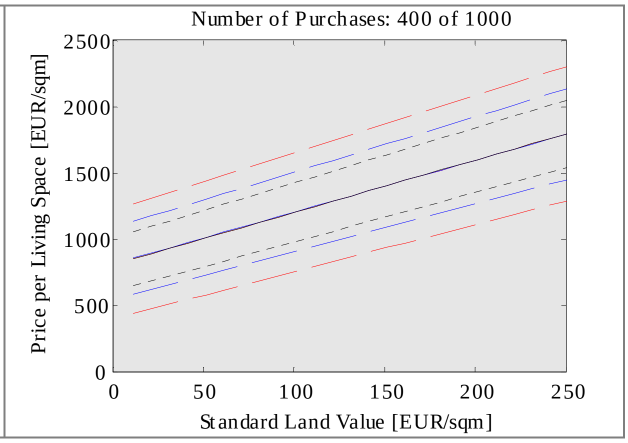

Fig. 4: Regression function for 400 purchases.

Similar to the general conclusion, the Bayesian approach improves the uncertainty, if small numbers of purchases are spread over the whole sample: Using informative prior, the confidence intervals are obviously smaller than the confidence intervals of the classic approach. The regression coefficients of the informative and the non-informative itself are nearly the same compared to the regression coefficients of the original sample. A few, but similar distributed purchases lead to a similar result as the whole sample.

4.2 Reduction to the center of the sample

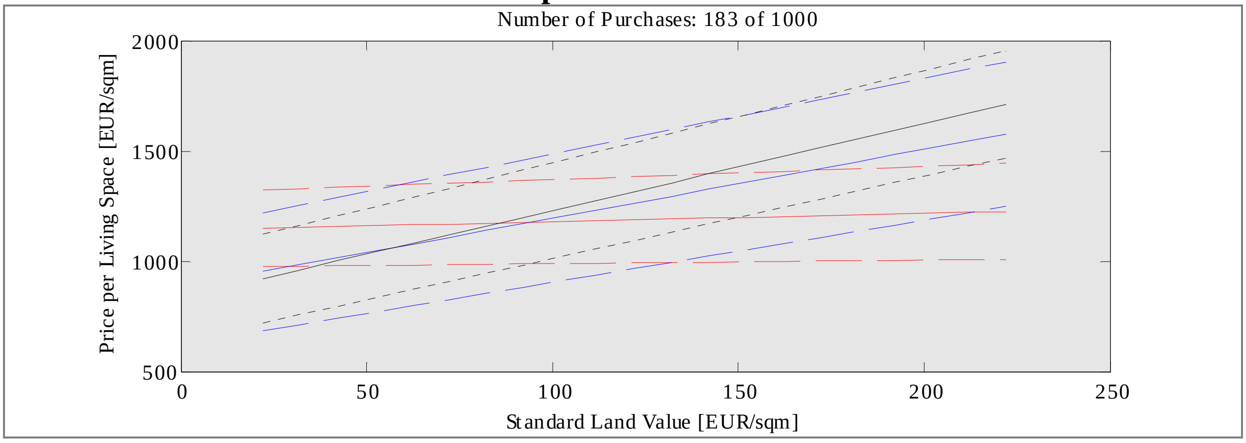

Fig. 5: Regression function for 183 purchases in the mean.

In a next step, the sample is reduced to its center. Therefore, the deciles are formed (the sample is divided in ten equal parts). Then the center is formed by cutting 10 % at each edge. Next steps are cutting 20 % to 40 % at each edge in steps of 10 %. Fig. 5 shows the result for the inner two deciles. The prices per living space in this sample varies from about 1100 to about 1300 €/m².

Key issue is, depending on whether informative and non-informative prior information is used, the regression functions differ very much. Comparing the reduced solutions with the original function, the informative prior information provides a more correct solution then the classical approach (red one). In Fig. 5, the regression function of the non-informative prior is only shown for standard land values between 70 and 150 €/m² within the confidence intervals of the regression function of the original sample.

Whereas the regression functions of the original sample and the informative prior increase for the variable standard land value, the regression function of the non-informative prior / classic approach is nearly constant. The results of the other parameters are correspondingly similar. The regression function of the informative prior differs from the original one only in the edges, but is completely within the confidence intervals of the original sample. Because of the lack of information in the edges, the slope of the function cannot be solved properly in the classical approach. An interpolation to the edges has to be done, which led to a very unstable solution. Here, the prior information can compensate absence of data in the edges.

As result of the cross validation, the difference between real prices of living space (therefore, 150 purchases are seperated before reduction) and estimated prices of living space varies, as well. The mean squarred error also differs. While the RMSE of informative prior is $RMSE_{inf}$ = 193 €/m², the RMSE of non-informative prior is nearly 68 €/m² higher ($RMSE_{non}$ = 261 €/m²).

4.3 Reduction to the edges of the sample

In a next step, the sample is reduced to its edges. Like in section 4.2, the edges are formed by deciles. In a first step, the outer both deciles are formed. Afterwards, each outer 20 % to 40 %

of the distribution forms the new sample. Fig. 6 presents the result for a reduction of the whole sample to the both outer deciles (at the edges).

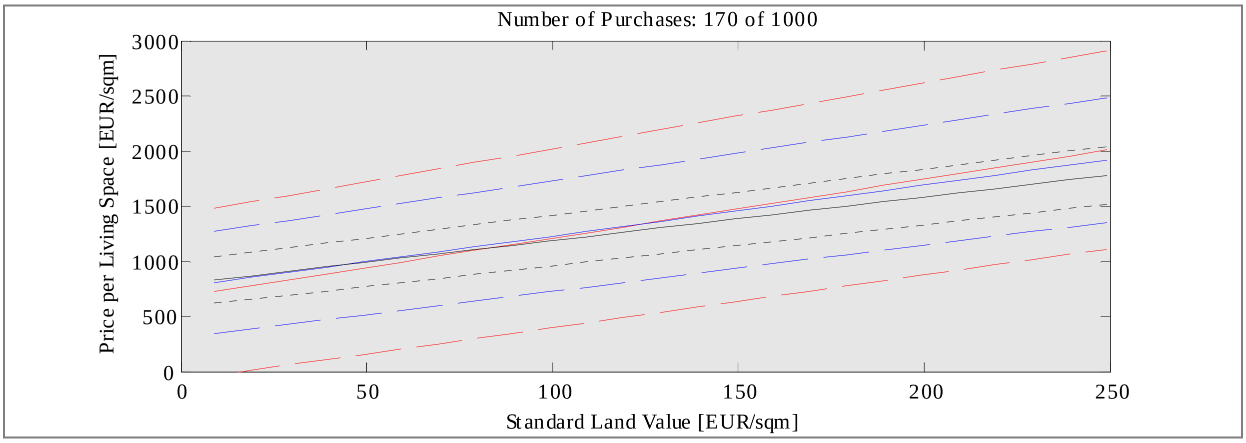

Fig. 6: Regression function for 170 purchases at the edges.

The prices per living space in this sample varies from about 300 to about 1000 €/m² and from about 1600 to about 2100 €/m².
Outcome of this analysis is a marginally better solution of the regression function using the informative prior information. But both solutions lie inside the confidence intervals of the original sample. A clear improvement is achieved for the confidence intervals. The confidence intervals of the informative prior information are narrower than the one of the classical approach. Although, there is a lack of information in the mean, the slope of the function is solved properly in the classical approach. Only a little compensation has to be done by the prior information. It is expected that this effect would be different, if functional relationship is not linear. In this case, the prior data would more important.
Besides, result of the cross-validation is a slightly better RMSE of the informative prior: $RMSE_{inf}$ = 181 €/m² according to $RMSE_{non}$ = 195 €/m². This means an improvement of nearly 14 €/m² by the informative prior information.

4.4 Reduction to the minimum edge of the sample

Thereafter, the sample is reduced to its minimum. The minimum 10 % to 40 % forms the new sample. Fig. 7 presents the solution for the minimum 2 deciles of the whole sample. The prices per living space in this sample varies from about 300 to about 1000 €/m².
Due to the lack of information about the higher purchasing prices, the regression function of the classical approach (with non-informative information) is only within the confidence intervals of the original data for standard land values smaller than 90 €/m².The same effect can be recognized like for a sample in the mean. The prior data compensate this lack of information. The regression function of the informative prior slightly differ from the original regression function. This applies not only to the parameter standard land value (showed in Fig. 7), but also for the other coefficients. While the classical solution has to interpolate, which led once more to a very unstable solution.

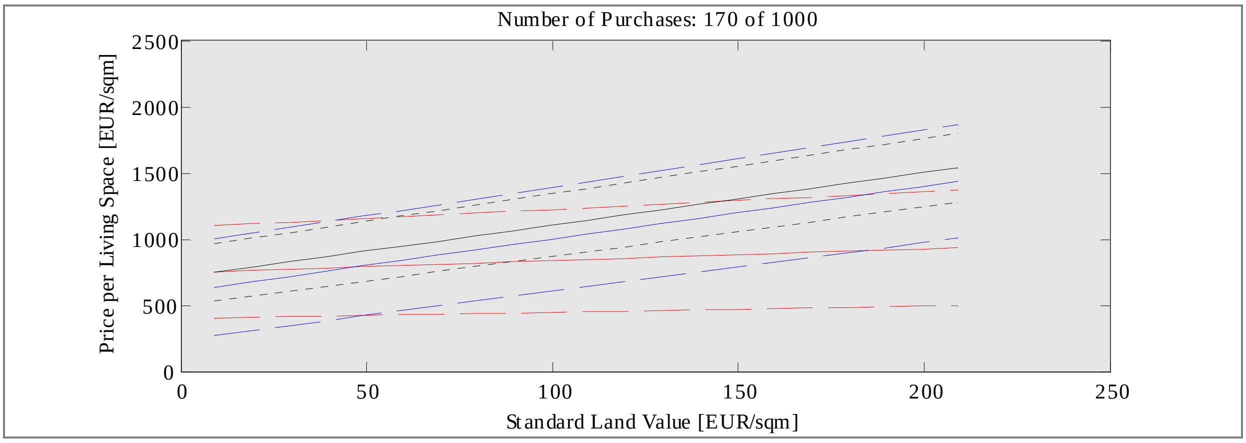

Fig. 7: Regression function for purchases at the minimum of the sample.

Besides, result of the cross-validation is a significantly better RMSE of the informative prior: $RMSE_{inf}$ = 262 €/m² versus $RMSE_{non}$ = 411 €/m². This means an improvement of almost 150 €/m² by the informative prior information.

4.5 Reduction to the maximum edge of the sample

As a last, the sample is reduced to its maximum. The maximum 10 % to 40 % forms the new sample. Fig. 8 illustrates the regression functions for the maximum 20 % of the original sample.

The prices per living space in this sample varies from about 1400 to about 2100 €/m². Although the regression function of the non-informative prior is within the confidence intervals of the original sample, but it is clear, that the slope is more different than the one of the informative prior. Here also the prior information provides an improvement over the non-informative solution. The regression function of the informative prior sightly diviates. This applies not only to the variable standard land value (showed in Fig. 8), but also for the other ones.

Once more, result of the cross-validation is a significantly better RMSE of the informative prior: $RMSE_{inf}$ = 239 €/m² versus $RMSE_{non}$ = 379 €/m². This means an improvement of almost 140 €/m² by the informative prior information.

Fig. 8: Regression function for purchases at the maximum of the sample.

5. DISCUSSION AND CONCLUSIONS

For randomized reduced samples, the regression coefficients of the informative and the non-informative are almost the same compared to the regression coefficients of the original sample. A few, but similar distributed purchases, lead to analogical results like using the whole sample.

The results of methodically reduced samples need a closer examination. The regression functions using informative prior information are more alike the ones of the original sample. A good improvement by using informative data is noticeable for the reduction to the center of the sample, to the minimum and to the maximum. This improvement arises whenever data is missing, which influences the slope. For this reason, only a slight improvement is remarkable for the reduction to the edges. In addition, it could be noticed that informative prior leads to an improvement for the reduction to the minimum and the maximum. Comparing the mean squared error, the best improvements by using informative prior are also remarkable for the reduction to the center of the sample. The same effect can be noticed for the reduction to the minimum and to the maximum (improvements of the RMSE 150 €/m² by using informative instead of non-informative prior). Once again, the improvement for the reduction to the edges is only a slight one.

Fig. 9 shows the regression coefficients (including the intercept) and the confidence intervals of the methodically reduced samples. In red shadowed color, the original confidence intervals are presented.The gray bars demonstrate the different confidence intervals of all the analysises (each informative and non-informative prior information). It should be noted that the confidence intervals are most of the solutions of informative prior lower than the solutions of non-informative prior. Only exception can be recognized for the reduction to the center of the sample. But the result is misleading. The estimation of the regression function is based on a center sample – through exactly this data is missing, which is needed to estimate a correct slope of the regression function. The confidence intervals seem more certain, because of the missing information. An interpretation can only be done for this trimmed data-set. The solution using informative prior is superior to the solution using non-informative prior comparing with the original data. Concerning the confidence intervals, the informative solution leads to more certain results for the rest of the analysis.

58 % of all confidence intervals of the informative solution overlap with the confidence interval of the original sample, while only 25 % of the non-informative solutions overlap. Good coincidences can be determined for the edges of the whole sample – using both informative and non-informative prior. Because of the linearity of the model, a lack of data from the mean can be compensated very well by the outer data. An obviously better solution is achieved by using informative prior in the analysis of mean, minimum or maximum data. Much more confidence intervals of the original and the reduced data overlap and much more the regression coefficients lie within the confidence intervals of the original data than with non-informative data. Overall, a quarter of the regression coefficients of the solution using informative prior lies within the confidence intervals of the original data, while only regression coefficient (sample of edges, area of lot) of the using non-informative prior lies within the confidence intervals of the original data.

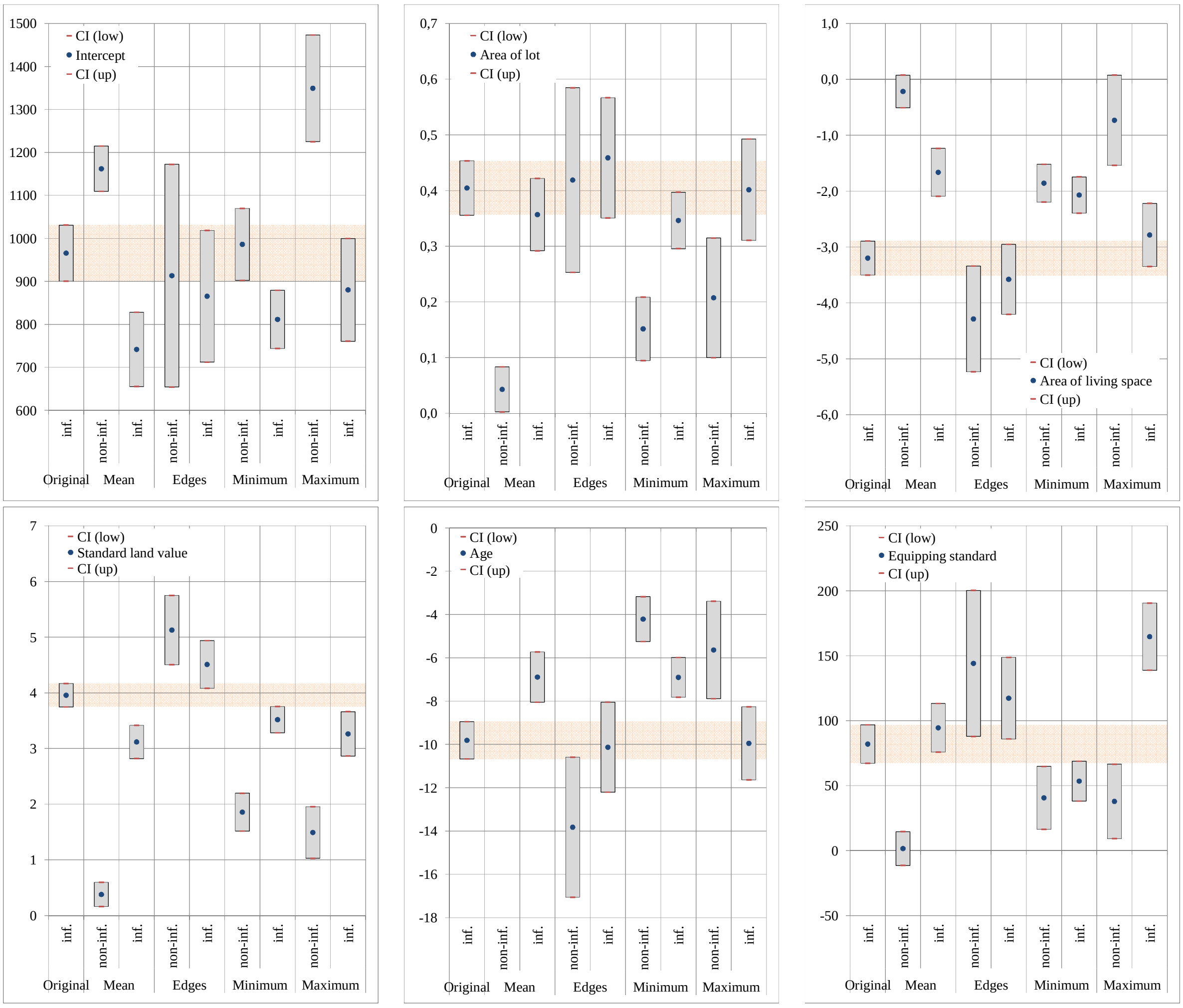

Fig. 9: Comparision of the different estimations of the regression coefficients and their confidence intervals.

As a clear statement, it should be noted that the solution using informative prior information leads to more certain estimation. It is an important remark, that the improvement of the regression function is particularly high for samples, where data is missing which influences the slope. As noted above, in case of low purchases, the market has often collapsed. Less good objects will not be sellable – these are the objects which would be found in the minima of a sample of purchases. If a transaction would be possible there, probably the better objects would sold. That means, that the data in the minimum is missing in such markets, so an estimation by classical regression analysis will lead to a worse slope. Also, the result would be more uncertain than by using informative prior information in a Bayesian approach.
The Bayesian approach is generally recommended. In “normal” market with sufficient data, it leads to more certain results. But the Bayesian approach achieved special relevance for the use in markets with few data. Here, in addition to a more certain result, the estimation itself is improved. But it still needs to be pointed out that the Bayesian approach demands good prior information. This information is obtained by the valuation experts, which must have a good qualification for the objective and spatial submarket. If this can be ensured, the use of the Bayesian approach is absolutely recommended

In further studies, an analysis should be applied in a ”real” market with low purchasing price. Also a spatial reference should be done. The particular challenge arises in the validation of the result. Unlike in this study, there is no comparison to a “whole” sample possible. A validation has to be developed, which controls the result by qualified and independent experts.

REFERENCES

Alkhatib, H., Weitkamp, A., 2012. Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung, Teil 1. In: ZfV 02.

Brückner, R., 1976. Mathematische Statistik bei der Ermittlung von Grundstückswerten - Lehrbriefe und Vorlesungen zum Kontaktstudium des Geodätischen Instituts 1976. Wissenschaftliche Arbeiten der Lehrstühle Geodäsie, Photogrammmetrie und Kartographie an der Technischen Hochschule Hannover. Hannover: Eigenverlag.

BVerfG, 2006. Urt. v. 07.11.2006 1 BvL 10/02.

Deutsches Institut für Normung, 2009. Allgemeine Anforderungen an Stellen, die Personen zertifizieren: General requirements for bodies operating certification of persons ; DIN EN ISO/IEC 17024. Berlin: Beuth 17024.

Fahrmeir, L., Kneib, T., Lang, S., 2009. Regression: Modelle, Methoden und Anwendungen. Heidelberg: Springer.

Gondring, H., 2004. Immobilienwirtschaft: Handbuch für Studium und Praxis. München: Vahlen.

Hartung, J., Elpelt, B., Klösener, K.-H., 2009. Statistik: Lehr- und Handbuch der angewandten Statistik ; [mit zahlreichen durchgerechneten Beispielen]. München: Oldenbourg.

Kacker, R., Jones, A., 2003. On use of Bayesian statistics to make the guide to the expression of uncertainty in measurement consistent. In: Metrologia 40 pp. 235–248.

Koch, K.-R., 1999. Parameter Estimation and Hypothesis Testing in Linear Models. Berlin - Heidelberg - New York: Springer.

Koch, K.-R., 2007. Introduction to Bayesian Statistics. Berlin - Heidelberg – New York: Springer.

Reuter, F., 2006. Zur Ermittlung von Bodenwerten in kaufpreisarmen Lagen. In: Flächenmanagement und Bodenordnung (03) pp. 97–107.

Sachs, L., Hedderich, J., 2006. Angewandte Statistik: Methodensammlung mit R ; mit 180 Tabellen. Berlin: Springer.

Urban, D., Mayerl, J., 2006. Regressionsanalyse: Theorie, Technik und Anwendung. Wiesbaden: VS Verl. für Sozialwiss.

Weitkamp, A., Alkhatib, H., 2012. Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung, Teil 2. In: ZfV (03) – accepted.

Willmott, C. J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. In: Climate Research 30 pp. 79–82.

Ziegenbein, W., 1977. Zur Anwendung multivarianter Verfahren in der mathematischen Statistik in der Grundstückswertermittlung. In: Wissenschaftliche Arbeiten der Lehrstühle Geodäsie, Photogrammetrie und Kartographie an der Technischen Hochschule Hannover.

Ziegenbein, W., 2009. Immobilienwertermittlung. In: Kummer, K., Frankenberger, J. Das deutsche Vermessungs- und Geoinformationswesen, pp. 423–471: Wichmann.

BIOGRAPHICAL NOTES

Dr. Alexandra Weitkamp received her Dipl.-Ing. in "Geodesy and Geoinformatics" at the University of Hanover in 1999. She passed the second state exam as "Graduate Civil Servant for Surveying and Real Estates" in Lower Saxony in 2001. After two-year experience at Bayer AG, she returns to Leibniz Universitaet Hannover. In 2008 received her Ph.D. in "Geodesy and Geoinformatics" at the University of Bonn. Since then, she has been postdoctoral fellow at the Geodetic Institute at the Leibniz Universitaet Hannover. Her main research interests are: adaption of innovative evaluation methods for valuation, stakeholders in rural and urban development, and decision-making methods.

Dr. Hamza Alkhatib received his Dipl.-Ing. in Geodesy and Geoinformatics at the University of Karlsruhe in 2001 and his Ph.D. in Geodesy and Geoinformatics at the University of Bonn in 2007. Since 2007 he has been postdoctoral fellow at the Geodetic Institute at the Leibniz Universitaet Hannover. His main research interests are: Bayesian Statistics, Monte Carlo Simulation, Modeling of Measurement Uncertainty, Filtering and Prediction in State Space Models, and Gravity Field Recovery via Satellite Geodesy.

CONTACTS

Dr.-Ing. Alexandra Weitkamp
Geodetic Insitute of
Leibniz Universitaet Hannover
– Land and Real Estate Management
Nienburger Str. 1
Hannover D - 30167
GERMANY
Tel. +49 (0) 511 762 2406
Fax +49 (0) 511 762 2468
Email weitkamp@gih.uni-hannover.de
Web www.gih.uni-hannover.de

Dr.-Ing. Hamza Alkhatib
Geodetic Insitute of
Leibniz Universitaet Hannover
– Evaluation methods
Nienburger Str. 1
Hannover D - 30167
GERMANY
Tel. +49 (0) 511 762 2464
Fax +49 (0) 511 762 2468
Email alkhatib@gih.uni-hannover.de
Web www.gih.uni-hannover.de